

# MIRACLE at ImageCLEFanoT 2007: Machine Learning Experiments on Medical Image Annotation

Sara Lana-Serrano<sup>1,3</sup>, Julio Villena-Román<sup>2,3</sup>  
José Carlos González-Cristóbal<sup>1,3</sup>, and José Miguel Goñi-Menoyo<sup>1</sup>

<sup>1</sup> Universidad Politécnica de Madrid

<sup>2</sup> Universidad Carlos III de Madrid

<sup>3</sup> DAEDALUS - Data, Decisions and Language, S.A.

slana@diatel.upm.es, jvillena@it.uc3m.es

josecarlos.gonzalez@upm.es, josemiguel.goni@upm.es

**Abstract.** This paper describes the participation of MIRACLE research consortium at the ImageCLEF Medical Image Annotation task of ImageCLEF 2007. Our areas of expertise do not include image analysis, thus we approach this task as a machine-learning problem, regardless of the domain. FIRE is used as a black-box algorithm to extract different groups of image features that are later used for training different classifiers based on kNN algorithm in order to predict the IRMA code. The main idea behind the definition of our experiments is to evaluate whether an axis-by-axis prediction is better than a prediction by pairs of axes or the complete code, or vice versa.

**Keywords:** Medical image, image annotation, classification, IRMA code, axis, learning algorithms, nearest-neighbour, machine learning.

## 1 Introduction

MIRACLE is a research consortium formed by research groups of three different universities in Madrid (Universidad Politécnica de Madrid, Universidad Autónoma de Madrid and Universidad Carlos III de Madrid) along with DAEDALUS, a private company founded as a spin-off of these groups and a leading company in the field of linguistic technologies in Spain. This paper describes our second participation [1] [2] in the ImageCLEF Medical Image Annotation task of ImageCLEF 2007 [3]. We approach this task as a machine learning problem, regardless of the domain, as our areas of expertise do not include image analysis research [4] and this task uses no textual information.

## 2 Description of Experiments

FIRE (Flexible Image Retrieval Engine) [5] [6] is a freely available content-based information retrieval system developed under the GNU General Public License that allows to perform query by example on images, using an image as the starting point for

the search process and relying entirely on the image contents. FIRE offers a wide repository of available features and distance functions. Specifically, the distribution package includes a set of scripts that extracts different types of features from the images [4], including color/gray histograms, invariant features histograms, Gabor features, global texture descriptor, Tamura features, etc.

Our approach to the task is to build different classifiers that use image features to predict the IRMA code [7]. For that purpose, all images in the training, development and testing dataset have been processed with FIRE. The extracted features have been divided into three groups, as shown in Table 1, to build the training data matrixes for the classifiers.

Table 1. Training data matrixes

| Name      | FIRE – Image Features   | Dimension <sup>1</sup> |
|-----------|---|------------------------|
| Histogram | Gray histogram and Tamura features  | 768                    |
| Vector    | Aspect ratio, global texture descriptor and Gabor features                                  | 75                     |
| Complete  | Gray histogram, Tamura features, aspect ratio, global texture descriptor and Gabor features | 843                    |

Different strategies have been evaluated, using several multiclassifiers built up with a set of specialized individual classifiers [2]:

- **IRMA Code Classifier:** single classifier that uses the image features to predict the complete IRMA code (4 axes: Technical, Direction, Anatomical and Biological).
- **IRMA Code Axis Classifier:** a two level classifier that is composed of four different classifiers that individually predict the value of each axis of the IRMA code; the prediction is the concatenation of partial solutions.
- **IRMA Code Combined Axis Classifier:** similar to the axis classifier, this one predicts the axes grouped in pairs.

All classifiers are based on the k-Nearest-Neighbour algorithm to predict the output class. After some preliminary runs, a value of k=10 was chosen. The main idea behind the definition of the experiments is to evaluate whether an axis-by-axis prediction is better than a prediction by pairs of axes or the complete code, or vice versa. In addition, the effect of applying the data normalization will be also analyzed. Table 2 shows an overview of the experiments. 30 experiments were finally submitted.

Table 2. Experiment set

| Features                      | Prediction <sup>2</sup>   | Normalization <sup>3</sup> |
|-------------------------------|---|----------------------------|
| Complete   Histogram   Vector | Complete code   Axis-by-axis   Combined axis: T+A and B+D   Combined axis: T+B and D+A   Combined axis: T+D and A+B | NO   YES                   |

<sup>1</sup> Number of columns of the matrix; the number of rows is 10,000 for the training dataset and 1,000 for the development and testing dataset.

<sup>2</sup> IRMA code axes are: Technical (**T**), Direction (**D**), Anatomical (**A**) and Biological (**B**).

<sup>3</sup> Normalized to range [0, 1].

### 3 Results

Results are shown in Table 3 [2]. According to the weighted error count score [8], which penalizes wrong decisions that are easy to take over wrong difficult decisions or at an early stage in the IRMA code, our best experiment is the one with data normalization that individually predicts each axis using all image features (“histogram” and “vector”). However, considering the number of correctly classified images, the best experiment is the one that uses normalized vector-based features and predicts the combined axis Technical+Direction and Anatomical+Biological.

**Table 3.** Evaluation of best-ranked experiments

| Run ID         | Error count   | Well classified |
|----------------|---------------|-----------------|
| MiracleAAn     | <b>158.82</b> | 497             |
| MiracleVAn     | 159.45        | 504             |
| MiracleAATDABn | 160.25        | 501             |
| MiracleAATABDn | 162.18        | 499             |
| MiracleVATDABn | 174.99        | <b>507</b>      |

On the other hand, comparing the predictions of the complete IRMA code versus the axis-by-axis predictions, the conclusion is that, regardless of the selected image features, the axis-by-axis prediction achieves more accurate results not only than the prediction of a combined pair of axes but also than the prediction of the complete code. It is interesting to observe that most groups have performed experiments focused on the prediction of the complete code.

In addition, data normalization seems to improve the predictions and vector-based features are preferred over histogram-based ones [2].

Our results were considerably worse, ompared to other groups’. The best experiment reached a score of 26.84, 17% of our own best error count. MIRACLE ranked 9<sup>th</sup> out of 10 participants in the task.

Probably different distance metrics should have been used to calculate the nearest neighbours. In particular, Mahalanobis distance, which is scale-invariant and takes into account the correlations among different variables, could have lead to better results.

However, we think that the main reason of the poor performance is the wrong choice of image features to train the classifiers. Although some feature selection experiments were carried out to reduce the high dimensions of the training data, no definite conclusion could be drawn and the complete set of features was finally used. Under these circumstances, the learning performance of the kNN algorithm is known to be worse than other algorithms’ such as SVM (Support Vector Machines), MLP (Multilayer Perceptrons) or Decision Trees. These classifiers will be considered for future participations in this task.

### 4 Conclusions and Future Work

The main conclusion that can be drawn from the evaluation is that, irrespective of the selected image features, the best experiments are those that predict the IRMA code

from the individual partial predictions of the 1-axis classifiers. Moreover, the predictions of combined pairs of axes are better than the predictions of the complete IRMA code. By extension, it could be concluded that the finer granularity of the classifier, the more accurate predictions are achieved. In the extreme case, the prediction may be built up from 13 classifiers, one per each character of the IRMA code. This issue will be further investigated and some experiments are already planned.

One of the toughest challenges to face when designing a classifier is the selection of the vector of features that best captures the different aspects that allow distinguishing one class from the others. Obviously, this requires an expert knowledge of the problem to be solved, which we currently lack. We are convinced that one of the weaknesses of our system is the feature selection. Therefore more effort will be invested in improving this topic for future participations.

**Acknowledgements.** This work has been partially supported by the Spanish R&D National Plan, by means of the project RIMMEL (Multilingual and Multimedia Information Retrieval, and its Evaluation), TIN2004-07588-C03-01; and by the Madrid's R&D Regional Plan, by means of the MAVIR project (Enhancing the Access and the Visibility of Networked Multilingual Information for the Community of Madrid), S-0505/TIC/000267.

## References

1. Villena-Román, J., González-Cristóbal, J.C., Goñi-Menoyo, J.M., Martínez Fernández, J.L.: MIRACLE's Naive Approach to Medical Images Annotation. In: Working Notes for the CLEF 2005 Workshop, Vienna, Austria (2005)
2. Lana-Serrano, S., Villena-Román, J., González-Cristóbal, J.C., Goñi-Menoyo, J.M.: MIRACLE at ImageCLEFnet 2007: Machine Learning Experiments on Medical Image Annotation. In: Working Notes for the CLEF 2007 Workshop, Budapest, Hungary (2007)
3. Deselaers, T., Müller, H., Deserno, T.: Automatic Medical Image Annotation in ImageCLEF 2007: Overview, Results, and Discussion. Pattern Recognition Letters, Special Issue on Medical Image Annotation in ImageCLEF 2007 (to appear, 2008)
4. Goodrum, A.A.: Image Information Retrieval: An Overview of Current Research. Informing Science 3(2), 63–66 (2000)
5. Deselaers, T., Keysers, D., Ney, H.: FIRE - Flexible Image Retrieval Engine: ImageCLEF 2004 Evaluation. In: Peters, C., Clough, P., Gonzalo, J., Jones, G.J.F., Kluck, M., Magnini, B. (eds.) CLEF 2004. LNCS, vol. 3491, pp. 688–698. Springer, Heidelberg (2005)
6. FIRE: Flexible Image Retrieval System, <http://www-i6.informatik.rwth-aachen.de/~deselaers/fire.html>
7. IRMA project: Image Retrieval in Medical Applications, <http://www.irma-project.org/>
8. Deselaers, T., Kalpathy-Cramer, J., Müller, H., Deserno, T.: Hierarchical classification for ImageCLEF, Medical Image Annotation (2007), <http://www-i6.informatik.rwth-aachen.de/~deselaers/imageclef07/hierarchical.pdf>